CLAIMS

1. A machine-based method comprising

in connection with a project in which a user generates a predictive model based on historical data about a system being modeled:

selecting variables having at least a predetermined level of significance from a pool of potential predictor variables associated with the data, to form a population of predictor variables,

extending the population to include non-linear interactions of variables,

extending the population to include linear and non-linear extensions with remaining previously excluded variables,

generating a possible model of the extended population of variables using a subsample of the data,

determining whether the model generalizes to the data other than the subsample,

if so, applying the possible model to all of the data to generate a final model, and cross-validating the final model using random portions of the data.

2. The method of claim 1 also including displaying information to the user of the size of the pool of predictor variables.

3. The method of claim 1 also including enabling a user to point and click to reduce or extend the size of the pool of predictor variables, retaining only the most significant predictor variables.

4. The method of claim 1 in which the user is enabled to invoke an automatic process to select a class of models most suitable to the pool of predictor variables for the associated dataset.

5. The method of claim 1 in which the user is enabled to point and click to adjust the model selection criterion to retain only the most potent variables for the target goal.

6. The method of claim 1 in which the user is enabled to point and click to cause display of information about the model performance.

7. The method of claim 6 in which the information about the model performance includes at least one of: a statistical report card with a link to the statistical report chart, a lift chart with a link to the lift chart, a response comparison chart for each decile for each predictor variable in the model, and a link to the response comparison chart.

8. The method of claim 7 in which invocation of the link to the statistical report card causes display of the statistics of model performance.

9. The method of claim 7 in which invocation of the link to the lift chart causes display of a non-cumulative lift chart.

10. The method of claim 7 in which invocation of the link to the response comparison chart causes display of a response chart for each predictor variable in the model for each segment of interest.

11. The method of claim 1 in which a user is enabled to choose interactively at least one performance criterion change or transformation or interaction of variables to improve the fit of the model.

12. The method of claim 1 also including a enabling a determination whether the model generalizes to the data other than the subsample, and, if so, applying the possible model to all of the data to generate a final model, and cross-validating the final model using random portions of the data.

13. The method of claim 12 in which the user is enabled to select at least one validation dataset and invoke a model process validation method.

14. The method of claim 12 in which the user is enabled to point and click to cause display of information about the model process validation.

15. The method of claim 12 in which the information about the model process validation includes at least one of: a statistical report card with a link to the statistical report chart, a cumulative lift chart with a link to the cumulative lift chart, a non-cumulative lift chart with a link to the non-cumulative lift chart.

16. The method of claim 1 in which the user is enabled to select at least one machine automated model development process applied to the entire dataset for a validated model process.

17. The method of claim 1 in which the user is enabled to point and click to cause display of information about the performance of the validated model process applied to the entire set of historical data.

18. The method of claim 17 in which the information about the model performance for two independent data subsets includes at least one of: a statistical report card with a

link to the statistical report chart, a cumulative lift chart with a link to the cumulative lift chart, a non-cumulative lift chart with a link to the non-cumulative lift chart.

19. The method of claim 18 in which the invocation of the link to the statistical report card causes display of the statistics of model process validation.

20. The method of claim 18 in which the invocation of the link to the cumulative lift chart causes display of a cumulative lift chart.

21. The method of claim 18 in which the invocation of the link to the cumulative lift chart causes display of a non-cumulative lift chart.

22. The method of claim 18 in which the final model and the model process validation results are stored persistently.

23. The method of claim 1 also including enabling the user to observe the number of predictor variables available for the model.

24. The method of claim 1 in which a model method from a library of model methods most suitable to modeling the historical data set is automatically selected.

25. The method of claim 1 also including enabling the user to observe the performance of the model by means of links to a plurality of statistical and graphical reports.

26. The method of claim 1 also enabling the user to select a means of validating the model development process.

27. The method of claim 1 also enabling the user to observe the performance of the model for the training and validation subsets of the historical dataset.

28. The method of claim 1 also enabling the user to invoke at least one validated model development process to produce a final model enabling the use to observe the performance of the final model on at least two independent subsets.

29. The method of claim 1 enabling the persisting of the final model and intermediate results to the project database.

30. The method of claim 1 enabling the final model to be applied to scoring at least one non-historical dataset wherein the propensity computed by the model is indexed by the score.

31. A machine-based method comprising

in connection with a project in which a user generates a predictive model based on historical data about a system being modeled, displaying to a user a lift chart, monotonicity, and concordance scores associated with each step in a step-wise model fitting process.

32. The method of claim 31 also including enabling the user to observe changes in the fit of the model as variables associated with the data are added or removed from a predictor set of the variables.

33. The method of claim 31 also including enabling the user to terminate the fitting of the model when the fitting process reaches an optimal point.

34. A machine-based method comprising

receiving from separate sources, sets of variables representing historical data about a system being modeled, and

enabling a user of a model generation tool to combine at least two of the variables from the sets of variables.

35. The method of claim 34 in which enabling the user to combine the variables includes providing a user interface that enables the user to identify the variables to be combined.

36. The method of claim 34 in which the system comprises behavior of prospective or current customers with respect to products or services of a company and the method also includes adjusting outcome variables to normalize response rates across products or services with different response rates.

37. The method of claim 34 in which the system comprises behavior of current customers with respect to retention of a current service or product of a vendor and the method also includes adjusting variables to normalize response rates across products or services with different response rates, using the computed propensities as indices of the scores.

38. The method of claim 31 also comprising determining a course of action to yield the most positive NPV outcome.

39. The method of claim 38 in which the determining includes selection of a mix of channel and product combinations.

40. The method of claim 38 in which the determining includes predicting retention in combination with response rate to predict NPV.